Overview of DrugProt BioCreative VII track: quality evaluation and large scale text mining of drug-gene/protein relations

Antonio Miranda[1], Farrokh Mehryary[2], Jouni Luoma[2], Sampo Pyysalo[2], Alfonso Valencia[1] and Martin Krallinger[1]
[1]Barcelona Supercomputing Center, Spain
[2]Turku University, Finland

Abstract

Considering recent progress in NLP, deep learning techniques and biomedical language models there is a pressing need to generate annotated resources and comparable evaluation scenarios that enable the development of advanced biomedical relation extraction systems that extract interactions between drugs/chemical entities and genes, proteins or miRNAs. Building on the results and experience of the CHEMDNER, CHEMDNER patents and ChemProt tracks, we have posed the DrugProt track at BioCreative VII. The DrugProt track focused on the evaluation of automatic systems able to extract 13 different types of drug-genes/protein relations of importance to understand gene regulatory and pharmacological mechanisms. The DrugProt track addressed regulatory associations (direct/indirect, activator/inhibitor relations), certain types of binding associations (antagonist and agonist relations) as well as metabolic associations (substrate or product relations). To promote development of novel tools and offer a comparative evaluation scenario we have released 61,775 manually annotated gene mentions, 65,561 chemical and drug mentions and a total of 24,526 relationships manually labeled by domain experts. A total of 30 teams submitted results for the DrugProt main track, while 9 teams submitted results for the large-scale text mining sub-track that required processing of over 2,3 million records. Teams obtained very competitive results, with predictions reaching f-measures of over 0.92 for some relation types (antagonist) and f-measures across all relation types close to 0.8.

INTRODUCTION

Among the most relevant biological and pharmacological relation types are those that involve (a) chemical compounds and drugs as well as (b) gene products including genes, proteins, miRNAs. A variety of associations between chemicals and genes/proteins are described in the biomedical literature, and there is a growing interest in facilitating a more systematic extraction of these relations from the literature, either for manual database curation initiatives or to generate large knowledge graphs of importance for drug discovery, drug repurposing, building regulatory or interaction networks or to characterize off-target interactions of drugs that might be of importance to understand better adverse drug reactions.

At BioCreative VI, the ChemProt track tried to promote the development of novel systems between chemicals and genes for groups of biologically related association types (ChemProt track relation groups or CPRs). Although the obtained results did have a considerable impact in the development and evaluation of new biomedical relation extraction systems, a limitation of grouping more specific relation types into broader groups was the difficulty to directly exploit the results for database curation efforts and biomedical knowledge graph mining application scenarios.

The considerable interest in the integration of chemical and biomedical data for drug-discovery purposes, together with the ongoing curation of relationships between biological and chemical entities from scientific publications and patents due to the recent COVID-19 pandemic, motivated the DrugProt track of BioCreative VII, which proposed using more granular relation types. In order to facilitate the development of more granular relation extraction systems large manually annotated corpora are needed. Those corpora should include high-quality manually labled entity mentions together with exhaustive relation annotations generated by domain experts.

TRACK AND CORPUS DESCRIPTION

Corpus description

To carry out the DrugProt track at BioCreative VII, we have released a large manually labelled corpus including annotations of mentions of chemical compounds and drugs as well as genes, proteins and miRNAs. Domain experts with experience in biomedical literature annotation and database curation annotated by hand all abstracts using the BRAT annotation interface. The manual labeling of chemicals and genes was done in separate steps and by different experts to avoid introducing biases during the text annotation process. The manual tagging of entity mentions of chemicals and drugs as well as genes, proteins and miRNAs was done following a carefully designed annotation process and in line with publicly released annotation guidelines. Gene/protein entity mentions

were manually mapped to their corresponding biologic al database identifiers whenever possible and classified as either *normalizable* to databases (tag: GENE-Y) or *non normalizable* mentions (GENE-N). Teams that participated at the DrugProt track were only provided with this classification of gene mentions and not the actual database identifier to avoid usage of external knowledge bases for producing their predictions.

The corpus construction process required first annotating exhaustively all chemical and gene mentions (phase 1). Afterwards the relation annotation phase followed (phase 2), were relationships between these two types of entities had to be labeled according to public available annotation guidelines. Thus, to facilitate the annotation of chemical-protein interactions, the DrugProt track organizers constructed very granular relation annotation rules described in a 33 pages annotation guidelines document. These guidelines were refined during an iterative process based on the annotation of sample documents.

The guidelines provided the basic details of the chemical-protein interaction annotation task and the conventions that had to be followed during the corpus construction process. They incorporated suggestions made by curators as well as observations of annotation inconsistencies encountered when comparing results from different human curators.

In brief, DrugProt interactions covered direct interactions (when a physical contact existed between a chemical/drug and a gene/protein) as well as indirect regulatory interactions that alter either the function or the quantity of the gene/gene product. The aim of the iterative manual annotation cycle was to improve the quality and consistency of the guidelines. During the planning of the guidelines some rules had to be reformulated to make them more explicit and clear and additional rules were added wherever necessary to better cover the practical annotation scenario and for being more complete.

The manual annotation task basically consisted of labeling or marking manually through a customized BRAT web-interface the interactions given the article abstracts as content. Figure 1 summarizes the DrugProt relation types included in the annotation guidelines.

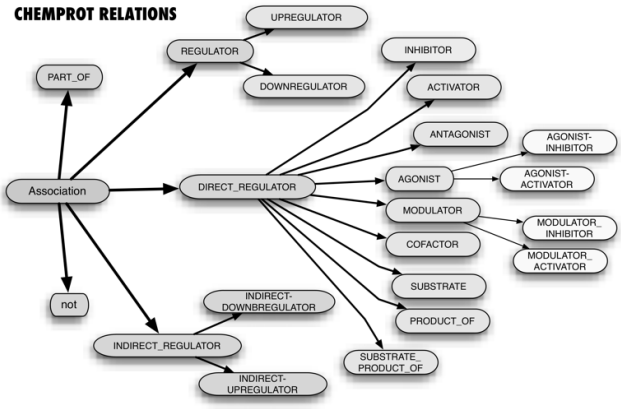

Fig. 1. Overview of the DrugProt relation type hierarchy.

The corpus annotation carried out for the DrugProt track was exhaustive for all the types of interactions previously specified. This implied that mentions of other kind of relationships between chemicals and genes (e.g. phenotypic and biological responses) were not manually labelled. Moreover, the DrugProt relations are directed in the sense that only relations of "what a chemical does to a gene/protein" (chemical → gene/protein direction) were annotated, and not vice versa.

To establish a easy to understand relation nomenclature and avoid redundant class definitions, we reviewed several chemical repositories that included chemical – biology information. We revised DrugBank, the Therapeutic Targets Database (TTD) and ChEMBL, assay normalization ontologies (BAO) and previously existing formalizations for the annotation of relationships: the Biological Expression Language (BEL), curation guidelines for transcription regulation interactions (DNA-binding transcription factor – target gene interaction) and SIGNOR, a database of causal relationships between biological entities.

Each of these resources inspired the definition of the subclasses DIRECT REGULATOR (e.g. DrugBank, ChEMBL, BAO and SIGNOR) and the INDIRECT REGULATOR (e.g. BEL, curation guidelines for transcription regulation interactions and SIGNOR). For example, DrugBank relationships for drugs included a total of 22 definitions, some of them overlapping with CHEMPROT subclasses (e.g. "Inhibitor", "Antagonist", "Agonist",…), some of them being regarded as highly specific for the purpose of this task (e.g. "intercalation", "cross-linking/alkylation") or referring to biological roles (e.g. "Antibody", "Incorporation into and Destabilization") and others, partially overlapping between them (e.g. "Binder" and "Ligand"), that were merged into a single class. Concerning indirect regulatory aspects, the five classes of casual relationships between a subject and an object term defined by BEL ("decreases", "directlyDecreases", "increases", "directlyIncreases" and "causesNoChange") were highly inspiring. Subclasses definitions of pharmacological modes of action were defined according to the UPHAR/BPS Guide to Pharmacology in 2016.

For the DrugProt track a very granular chemical-protein relation annotation was carried out, with the aim to cover most of the relations that are of importance from the point of view of biochemical and pharmacological/biomedical perspective. Nevertheless, for the DrugProt track only a total of 13 relation types were used, keeping those that had enough training instances/examples and sufficient manual annotation consistency. The final list of relation types used for this shared task was: INDIRECT-DOWNREGULATOR, INDIRECT-UPREGULATOR, DIRECT-REGULATOR, ACTIVATOR, INHIBITOR, AGONIST, ANTAGONIST, AGONIST-ACTIVATOR, AGONIST-INHIBITOR, PRODUCT-OF, SUBSTRATE, SUBSTRATE_PRODUCT-OF or PART-OF. The DrugProt corpus was split randomly into training, development and test set. We also included a background and large scale background collection of records that were automatically annotated with drugs/chemicals and genes/proteins/miRNAs using an entity tagger trained on the manual DrugProt entity mentions. The background collections were merged with the test set to be able to get team predictions also for these records. Table 1 shows a summary of the

DrugProt corpus in terms of number of entity annotations as well as relation annotations across each of the corpus subsets, while table 2 provides a more granular overview of the annotations for each of the relation classes used for the DrugProt track evaluation.

TABLE I. DRUGPROT CORPUS OVERVIEW

Set	Number of abstracts	Number of entities		Number of relations
		GENE	CHEMICAL	
Training	3500	43255	46274	17274
Development	750	9005	9853	3761
Test	750	9515	9434	3491
Background	10000	157523	134333	-
Large Scale	2366081	33578479	20415123	-

TABLE II. DETAILED DRUGPROT RELATIONS

Relation type	Nr. relations		
	Training	Development	Test
ANTAGONIST	1428	246	334
AGONIST	658	131	101
AGONIST-INHIBITOR	29	10	0
DIRECT-REGULATOR	13	2	3
ACTIVATOR	972	218	154
INHIBITOR	2247	458	429
INDIRECT-DOWNREGULATOR	1329	332	304
INDIRECT-UPREGULATOR	1378	302	277
PART-OF	5388	1150	1051
PRODUCT-OF	885	257	228
SUBSTRATE	920	158	181
SUBSTRATE_PRODUCT-OF	2003	494	419

Track participants had to return for the collection of test set document identifiers the detected pairs of entities (one corresponding to a chemical entity and another to a gene/protein) together with the corresponding relation type. Only relations between a chemical and a gene/protein were allowed. Relations between a chemical and another chemical or between a gene/protein and another gene/protein were not permitted. Moreover, participants were allowed to return for a given entity pair multiple relation groups. A total of 5 runs were accepted per team.

Evaluation

In addition to the DrugProt track data sets, a special evaluation script was available at the track webpage. For evaluation purposes we considered the micro-averaged precision, recall and balanced micro F1-score.

RESULTS

A total of 30 teams returned overall 107 submission runs for the Main subtrack of DrugProt. And 9 teams submitted 21 runs for the Large Scale subtrack of DrugProt. Table III lists the competing teams together with the results of the best run of each sub-track. A detailed description of the underlying strategy used by each of the participating teams can be found in the systems description papers published in the BioCreative VII workshop proceedings. The best F-measure, across all relations, was reached by team Humboldt (run 1) with a micro f1score of 79.73. And the highest micro f1-score of the Large Scale subtract was reached by team NLM-NNCBI (run 5) with 78.86. It is noteworthy that, for the Large Scale subtrack, the total number of relations extracted is 146M.

Performance varied depending on the particular class of chemical-protein relations. Table IV lists the runs with the highest precision, recall and f1-score for each relation type in the Main sub-track. For all relation types, except SUBSTRATE_PRODUCT-OF and AGONIST-INHIBITOR, team DigiLab-UG reached the highest recall scores. This is in line with the huge number of relations predicted by DigiLab-UG systems. Their submissions had an average of 130K predictions, while the average number of predictions in the Main Track was nearly 64K. The submission with the highest f1-score was reached by team NLM-NCBI (run 1) on the ANTAGONIST relation type: 92.99.

A more detailed view of the results per relation type is included in tables V, VI, VII and VIII. They contain the best submission run results obtained for each of the participating teams per relation type. The relations with the highest metrics are ANTAGONIST, AGONIST, INHIBITOR, ACTIVATOR and INDIRECT-DOWNREGULATOR. Indeed, for the ANTAGONIST relation type, there are 18 teams obtaining f1-scores larger than 85.0 in at least one of their submissions. For such relations, the number of training, development and test examples is considerably large. On the other hand, relation types such as PRODUCT-OF, SUBSTRATE, SUBSTRATE_PRODUCT-OF and AGONIST-INHIBITOR seem to be more complicated to detect for participating systems. Interestingly, PRODUCT-OF and SUBSTRATE have large support. There exists therefore an intrinsic difficulty of those two relation types.

Methodology

Two NLP trends are ubiquitous in the DrugProt participants. First, the usage of Transformer Language models, that have overcome the previous technologies (e.g. word

embeddings). Participants were about the language representations they employed and 30 out of 32 responses reported using transformer LM, while only 5 of them included also word embeddings. Together with the implementation of transformer LMs in NLP systems, there has been an explosion of the variety and specialization of such models. Figure 3 shows the reported ones that DrugProt participants have employed.

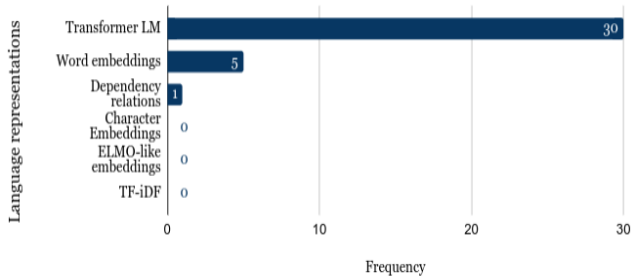

Fig. 2. Language representations used by DrugProt participants. Total number of participants: 32

Secondly, most participants, and particularly those with the highest performances report using ensemble systems. In most cases, such ensembles are simply the combination of the same system trained with different initializations, or with minor modifications. From the 19 teams included in the proceedings, 13 of them report using an ensemble system.

DrugProt teams have approached the relation extraction challenge from two approaches. Most teams modeled the challenge as a text (or sentence) classification task. The systems developed first divide the input text into fragments and then, each fragment is classified into a category. However, a small group of teams (such as NLM-NCBI and USMBA_UIT), modeled the problem as a Named Entity Recognition one. In this scenario, the system first divides the input text into fragments, and then, classifies every token of the fragment into a category. In both approaches, the categories are related to the listed relation types or to the null category. It is noteworthy that NLM-NCBI obtained its best results ensembling both approaches.

From the pre-processing part, there is a divergence among participants on how to treat the marked chemical and protein entities. There are three main options: entity masking (substitute them predefined tokens), entity marking (add markers indicating the beginning and end of the entities) or doing nothing. While most participants opted for the second strategy (entity marking) there has not been an extensive comparison of the strategies among the DrugProt participants.

For the language encoding, almost every team employed some pre-trained transformer language model. The decoding (classification) part, on the other hand, offers a wider variety. Most teams applied a simple linear classification layer (for example, Humboldt or FSU2021) or a softmax (such as NLM-NCBI and NLPatVCU) to the CLS token of the language model. However, some participants opted for more sophisticated approaches. For example, BIT.UA applied a multi-head attention mechanism, and CU-UD employed also LSTM.

Such innovations are relevant and show the wide range of tools available for solving the task. However, teams with the highest f1-scores employed rather simplistic decoding (or classification) mechanisms. They opted to invest their efforts in (A) ensembling many models and (B) enriching the encoding part of the system, using either external resources (Humboldt) data augmentation (NLM-NCBI) or combining both strategies (KU-AZ).

The system with the highest micro-f1 score was Humboldt. This team obtained as well the highest f1-score for the relation types DIRECT-REGULATOR, INDIRECT-UPREGULATOR, INHIBITOR and PRODUCT-OF. The authors defined the challenge as a sentence classification problem. Sentence were input to the biomedical pretrained transformer language models RoBERTa-large-PM-M3-Voc (30). The classification was performed with a linear layer applied to the CLS token embedding of the language model. Entity descriptions from the CTD database were used to enrich the model information. The best results were obtained ensembling ten models by averaging the predicted probabilities of every instance.

The NLM-NCBI team obtained the second-highest micro-f1 score and the highest f1-score for the relations ANTAGONIST, AGONIST, AGONIST-INHIBITOR, SUBSTRATE and PART_OF. They tested two separate approaches for solving the challenge: text classification and sequence labeling. Again, biomedical pre-trained language models are used for both frameworks including, but not only, PubMedBERT. On top of the LM, a softmax layer was applied on the CLS token output to perform text classification, while for the sequence labeling approach, predictions for each token were obtained applying a fully connected layer and a softmax classification layer. The best results were obtained ensembling with the "majority voting" strategy all the text classification and sequence labeling models.

Finally, the team KU-AZ obtained the third-highest micro-f1 score and the highest f1-score for the relation INDIRECT-DOWNREGULATOR and AGONIST-INHIBITOR. They augmented the DrugProt dataset by predicting labels with transformer models and built a larger dataset, that was refined with a knowledge base. Then, the challenge was modeled as a text classification task. Instances were passed through a biomedical pre-trained language model and a linear classification layer was applied on the embedding of the CLS token. Finally, models were ensembled. The authors report that data augmentation has worked particularly well for relation types with a low number of examples.

DISCUSSION

The DrugProt track attracted considerable interest by the biomedical text mining community, with over 100 registered teams and 30 submitting results. In addition to excellent results, close to human annotation quality for some of the relation types despite the underlying complexity. It is noteworthy to mention that some systems did also scale very well, being able to process over 2 million PubMed abstracts. DrugProt posed the first large scale biomedical text mining

task so far, generating in a collaborative way a very large biomedical knowledge graph, of high value for graph mining as well as biocuration initiatives.

The DrugProt corpus released for this track is also the largest of its kind so far released, both in terms of manually labeled entity mentions (over 120 thousand) as well as in terms of the number of manually annotated biomedical relations.

ACKNOWLEDGMENT *(Heading 5)*

DrugProt was promoted through the collaboration between the Spanish Plan for the Advancement of Language Technology (Plan TL) and the BSC. We also want to acknowledge the 2020 Proyectos de I+D+i - RTI Tipo A (PID2020-119266RA-I00)) for support.

REFERENCES

1. Leon Weber, Mario Sänger, Samuele Garda, Fabio Barth, Christoph Alt and Ulf Leser (2021) Humboldt @ DrugProt: Chemical-Protein Relation Extraction with Pretrained Transformers and Entity Descriptions. Proceedings of the seventh BioCreative challenge evaluation workshop.

2. Ling Luo, Po-Ting Lai, Chih-Hsuan Wei and Zhiyong Lu (2021) Extracting Drug-Protein Interaction using an Ensemble of Biomedical Pre-trained Language Models through Sequence Labeling and Text Classification Techniques. Proceedings of the seventh BioCreative challenge evaluation workshop.

3. Wonjin Yoon, Sean Yi, Richard Jackson, Hyunjae Kim, Sunkyu Kim and Jaewoo Kang (2021) Using Knowledge Base to Refine Data Augmentation for Biomedical Relation Extraction. KU-AZ team at the BioCreative 7 DrugProt challenge. Proceedings of the seventh BioCreative challenge evaluation workshop.

4. Avisha Das, Zhao Li, Qiang Wei, Jianfu Li, Liang-Chin Huang, Yan Hu, Rongbin Li, W. Jim Zheng and Hua Xu (2021) UTHealth@BioCreativeVII: Domain-specific Transformer Models for Drug-Protein Relation Extraction. Proceedings of the seventh BioCreative challenge evaluation workshop.

5. Anfu Tang, Louise Deléger, Robert Bossy, Pierre Zweigenbaum and Claire Nédellec (2021) Does constituency analysis enhance domain-specific pre-trained BERT models for relation extraction? Proceedings of the seventh BioCreative challenge evaluation workshop.

6. Mehmet Efruz Karabulut, K. Vijay-Shanker and Yifan Peng (2021) CU-UD: text-mining drug and chemical-protein interactions with ensembles of BERT-based models. Proceedings of the seventh BioCreative challenge evaluation workshop.

7. Naoki Iinuma, Masaki Asada, Makoto Miwa and Yutaka Sasaki (2021) TTI-COIN at BioCreative VII Track 1. Drug-protein interaction extraction with external database information. Proceedings of the seventh BioCreative challenge evaluation workshop.

8. Xin Sui, Wanjing Wang and Jinfeng Zhang (2021) Text Mining Drug-Protein Interactions using an Ensemble of BERT, Sentence BERT and T5 models. Proceedings of the seventh BioCreative challenge evaluation workshop.

9. Virginia Adams, Hoo-Chang Shin, Carol Anderson, Bo Liu and Anas Abidin (2021) Text Mining Drug/Chemical-Protein Interactions using an Ensemble of BERT and T5 Based Models. Proceedings of the seventh BioCreative challenge evaluation workshop.

10. Qingyao Li, Ying Xiong, Jingyue Hu and Buzhou Tang (2021) Using Knowledge-Based Pretrained Language Model for Mining Drug and Chemical-Protein Interactions. Proceedings of the seventh BioCreative challenge evaluation workshop.

11. Ed-drissiya El-allaly, Mourad Sarrouti, Noureddine En-Nahnahi and Said Ouatik El Alaoui (2021) A Multi-Task Transfer Learning-based method for extracting Drug-Protein Interactions. Proceedings of the seventh BioCreative challenge evaluation workshop.

12. Darshini Mahendran, Sudhanshu Ranjan, Jiawei Tang, Mai H. Nguyen, and Bridget T. McInnes (2021) BioCreative VII-Track 1: A BERT-based System for Relation Extraction in Biomedical Text. Proceedings of the seventh BioCreative challenge evaluation workshop.

13. Rui Antunes, Tiago Almeida, João Figueira Silva and Sérgio Matos (2021) Chemical–protein relation extraction in PubMed abstracts using BERT and neural networks. Proceedings of the seventh BioCreative challenge evaluation workshop.

14. Parsa Bagherzadeh and Sabine Bergler (2021) Dependencies for Drug-Prot Relation Extraction CLaC at BioCreative VII Track 1. Proceedings of the seventh BioCreative challenge evaluation workshop.

15. Dennis N. Mehay and Kuan-Fu Ding (2021) Catalytic DS at BioCreative VII: DrugProt Track A Syntactically-guided BiLSTM with BERT-derived Word Vectors. Proceedings of the seventh BioCreative challenge evaluation workshop.

16. Jenny Copara and Douglas Teodoro (2021) Drug-protein relation extraction using ensemble of transformer-based language models. Proceedings of the seventh BioCreative challenge evaluation workshop.

17. Diana Sousa, Rodrigo Cassanheira and Francisco M. Couto (2021) lasigeBioTM at BioCreative VII Track 1: Text mining drug and chemical-protein interactions using biomedical ontologies*. Proceedings of the seventh BioCreative challenge evaluation workshop.

18. Ting-Wei Chang, Tzu-Yi Li, Yu-Wen Chiu, Sheng-Jie Lin, Panchanit Boonyarat, Wen-Chao Yeh, Neha Warikoo and Yung-Chun Chang (2021) Identifying Drug/chemical-protein Interactions in Biomedical Literature using the BERT-based Ensemble Learning Approach for the BioCreative 2021 DrugProt Track. Proceedings of the seventh BioCreative challenge evaluation workshop.

19. Jehad Aldahdooh, Ziaurrehman Tanoli and Jing Tang (2021) R-BERT-CNN: Drug-target interactions extraction from biomedical literature BioCreative Challenge VII Track 1 submission. Proceedings of the seventh BioCreative challenge evaluation workshop.

20. Humboldt (2021), drugprot, https://github.com/leonweber/drugprot

21. FSU2021 (2021), ChemProt-BioCreative, https://github.com/luckynozomi/ChemProt-BioCreative

22. bibliome (2021), drugprot-relation-extraction, https://github.com/Maple177/drugprot-relation-extraction

23. NVHealthNLP (2021), Relation_Extraction-BioMegatron, https://github.com/NVIDIA/NeMo/blob/main/tutorials/nlp/Relation_Extraction-BioMegatron.ipynb

24. CU-UD (2021), drugprot_bcvii, https://github.com/bionlplab/drugprot_bcvii

25. USMBA_UIT (2021), mttl-drugprot, https://github.com/drissiya/mttl-drugprot

26. lasigeBioTM (2021), biocreativeVII, https://github.com/lasigeBioTM/biocreativeVII

27. NLPatVCU (2021), BioCreative-VII-Track1, https://github.com/NLPatVCU/BioCreative-VII-Track1

28. Orpailleur (2021), Relation-Extraction---DrugProt, https://github.com/FatimaHabib/Relation-Extraction---DrugProt

29. Jungfraujoch (2021), chemprot-drugprot_testing_ground, https://github.com/wenyuan-wu/chemprot-drugprot_testing_ground

30. Lewis, P., Ott, M., Du, J., & Stoyanov, V. (2020, November). Pretrained Language Models for Biomedical and Clinical Tasks: Understanding and Extending the State-of-the-Art. In Proceedings of the 3rd Clinical Natural Language Processing Workshop (pp. 146-157).

APPENDIX

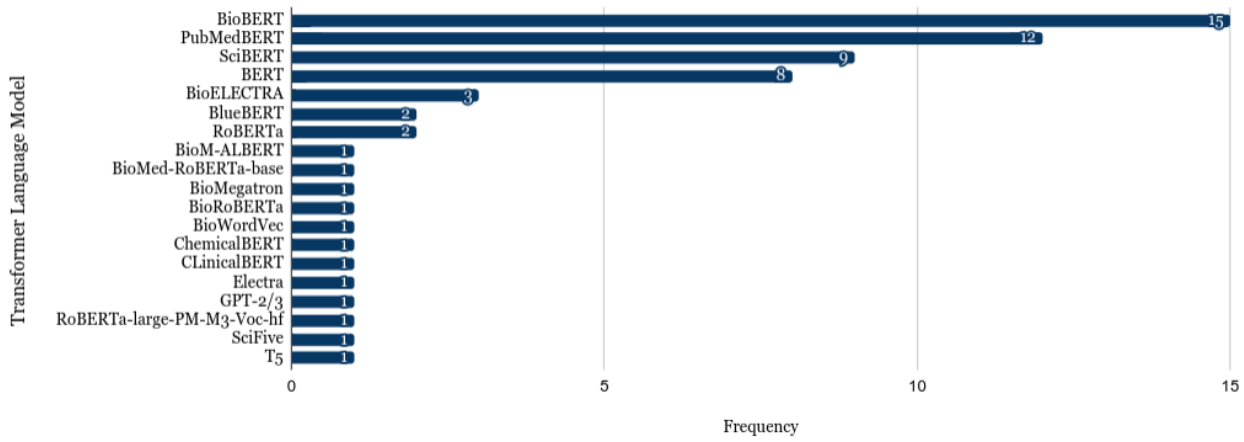

Fig. 3. Transformer Language Models used by DrugProt participants.

TABLE III. TEAM OVERVIEW, MAIN TRACK, AND LARGE SCALE MICRO-AVERAGE RESULTS

#	Team	*Affiliation*	*Ref*	*Tool URL*	*Main Track*				*Large Scale Track*			
					P	***R***	***F1***	***run***	***P***	***R***	***F1***	***run***
15	Humboldt	Humboldt-Universität Berlin, Germany	1	20	0.7961	0.7986	0.7973	1				
18	NLM-NCBI	National Institutes of Health, USA	2		0.7847	0.8052	0.7948	5	0.7730	0.8049	0.7886	***2***
13	KU-AZ	Korea University, AstraZeneca, AIGEN Sciences, South Korea, UK	3		0.7972	0.7817	0.7894	2	0.7644	0.7521	0.7582	2
7	UTHealth-CCB	University of Texas, USA	4		0.8044	0.7496	0.7760	2	0.7949	0.7527	0.7732	3
21	bibliome	INRAE France	5	22	0.7546	0.7966	0.7750	2				
3	CU-UD	University of Delaware, USA	6	24	0.7709	0.7771	0.7740	3	0.7466	0.7808	0.7633	1
29	TTI-COIN	Toyota Technological Institute, Japan	7		0.74931	0.7777	0.7632	1				
4	good team	Guangdong University of Foreign Studies, China	-		0.7344	0.7940	0.7630	5	0.7201	0.766762	0.7427	1
23	FSU2021	Florida State University, USA	8	21	0.7540	0.7510	0.7525	4	0.7066	0.7272	0.7167	1
14	HY-NLP	Hanyang University, South Korea	-		0.7122	0.7920	0.7500	1				
28	NVhealthNLP	NVIDIA, USA	9	23	0.7732	0.7249	0.7483	4	0.7325	0.332665	0.4575	1
16	HITSZ-ICRC	Harbin Institute of Technology, China	10		0.7671	0.7183	0.7419	4				
6	Saama Research	Saama Technologies, India	-		0.7406	0.7361	0.7383	1				
10	Stelios	-, Greece	-		0.7315	0.7261	0.7288	4				
5	The Three Musketeers	Fudan University, China	-		0.6993	0.7564	0.7268	1	0.6937	0.5860	0.635	1
2	USMBA_UIT	Sidi Mohamed Ben Abdellah University, Morocco	11	25	0.7569	0.6745	0.7133	4				

19	NLPatVCU	Virginia Commonwealth University, USA	12	27	0.7335	0.6908	0.7115	1				
27	BIT.UA	University of Aveiro, Portugal	13		0.7003	0.7229	0.7114	2				
25	Jungfraujoch	University of Zurich & ETH Zurich, Switzerland	-	29	0.7798	0.6201	0.6908	1				
24	CLaC	Concordia University, Canada	14		0.6444	0.7014	0.6717	3				
26	catalytic	Catalytic DS, Inc., United States	15		0.6746	0.5822	0.6250	1				
8	DigiLab-UG	University of Geneva, Switzerland	16		0.4507	0.8794	0.5959	4				
1	Trerotola	University of Brescia, Italy	-		0.3149	0.8378	0.4578	1				
17	BHAM	University of Birmingham, UK	-		0.2305	0.3673	0.2833	1				
11	LasigeBioTM	LASIGE, Portugal	17	26	0.3690	0.1865	0.2478	1				
9	TMU_NLP	Taipei Medical University, Taiwan	18		0.5678	0.1224	0.2013	2	0.4502	0.8287	0.5834	2
12	Elsevier Health Data Science	Elsevier, USA	-		0.5947	0.0576	0.1050	1				
20	Orpailleur	Université de Lorraine, CNRS, France	-	28	0.3078	0.0438	0.0767	3				
30	NetPharMed	University of Helsinki, Finland	19		0.0395	0.1573	0.0631	1				
22	CanSa	Al Baha University, Saudi Arabia	-		0	0	0	1				
mean					0.6430	0.6430	0.6430		0.7136	0.7446	0.7166	
std					0.1962	0.2472	0.2317		0.7120	0.7428	0.7142	
maximum					0.8044	0.879	0.7973		0.7949	0.8287	0.7886	

TABLE IV. BEST PRECISION, BEST RECALL AND BEST F1SCORE RUNS PER RELATION TYPE

Team	***Highest-precision run***				***Highest-recall run***				***Highest-f1score run***			
	P	***R***	***F1***	***run***	***P***	***R***	***F1***	***run***	***P***	***R***	***F1***	***run***
ANTAGONIST	0.9167	0.9346	0.9256	KU-AZ-1	0.5896	0.9673	0.7327	DigiLab-UG-3	0.9068	0.9542	0.9299	NLM-NCBI-1
AGONIST	0.8977	0.7822	0.8360	KU-AZ-4	0.5131	0.9703	0.6712	DigiLab-UG-4	0.8830	0.8218	0.8513	NLM-NCBI-2
AGONIST-INHIBITOR	1	1	1	many	1	1	1	many	1	1	1	many
DIRECT-REGULATOR	0.7753	0.6434	0.7032	Humboldt-3	0.3753	0.8485	0.5204	DigiLab-UG-4	0.7582	0.7016	0.7288	Humboldt-1
INHIBITOR	0.8953	0.8620	0.8783	Humboldt-3	0.55	0.9001	0.6828	DigiLab-UG-3	0.8801	0.8801	0.8801	Humboldt-1
ACTIVATOR	0.8641	0.7994	0.8305	NLM-NCBI-1	0.4855	0.9042	0.6318	DigiLab-UG-4	0.8440	0.8263	0.8351	bibliome-5
PRODUCT-OF	0.7464	0.5691	0.6458	Jungfraujoch	0.2791	0.8343	0.4183	DigiLab-UG-3	0.6733	0.754	0.7102	Humboldt-1
SUBSTRATE	0.7825	0.6611	0.7167	NLM-NCBI-4	0.3397	0.8449	0.4846	DigiLab-UG-3	0.7721	0.6874	0.7273	NLM-NCBI-3
SUBSTRATE_PRODUCT-OF	1	0.1	0.1818	bibliome-2	1	0.1	0.1818	bibliome-2	1	0.1	0.1818	bibliome-2
INDIRECT-DOWNREGULATOR	0.7880	0.4770	0.5943	Catalytic-2	0.4857	0.8947	0.6296	DigiLab-UG-2	0.7588	0.8487	0.8012	KU-AZ-2
INDIRECT-UPREGULATOR	0.8163	0.2888	0.4267	Catalytic-2	0.4509	0.8953	0.5998	DigiLab-UG-4	0.7770	0.7798	0.7784	Humboldt-3
PART_OF	1	0.0044	0.0087	LasigeBioTM-4	0.4174	0.8860	0.5674	DigiLab-UG-3	0.7531	0.8026	0.778	NLM-NCBI-2

TABLE V. MAIN TRACK GRANULAR RESULTS I

#	Team	*ANTAGONIST*				*AGONIST*				*AGONIST-INHIBITOR*			
		P	***R***	***F1***	***run***	***P***	***R***	***F1***	***run***	***P***	***R***	***F1***	***run***
15	Humboldt	0.889571	0.947712	0.917722	3	0.803738	0.851485	0.826923	4	1	0.333333	0.5	5
18	NLM-NCBI	0.906832	0.954248	0.929936	1	0.882979	0.821782	0.851282	2	1	1	1	all
13	KU-AZ	0.916667	0.934641	0.925566	1	0.897727	0.782178	0.835979	3,4	1	1	1	all
7	UTHealth-CCB	0.872727	0.941176	0.90566	5	0.824176	0.742574	0.78125	4	1	1	1	2:5
21	bibliome	0.895062	0.947712	0.920635	1	0.851064	0.792079	0.820513	2	0	0	0	all
3	CU-UD	0.878788	0.947712	0.91195	5	0.763636	0.831683	0.796209	1	0	0	0	all
29	TTI-COIN	0.861446	0.934641	0.896552	5	0.78	0.772277	0.776119	1	1	1	1	5
4	good team	0.902597	0.908497	0.905537	2	0.808081	0.792079	0.8	5	1	1	1	2
23	FSU2021	0.86875	0.908497	0.888179	4	0.724138	0.831683	0.774194	2	0	0	0	all
14	HY-NLP	0.872727	0.941176	0.90566	2	0.813725	0.821782	0.817734	1	1	1	1	all
28	NVhealthNLP	0.849398	0.921569	0.884013	4	0.761905	0.792079	0.776699	3	1	0.666667	0.8	3
16	HITSZ-ICRC	0.883562	0.843137	0.862876	4	0.822222	0.732673	0.774869	4	1	1	1	2,3
6	Saama Research	0.865385	0.882353	0.873786	2	0.765306	0.742574	0.753769	2	1	1	1	1,3
10	Stelios	0.878378	0.849673	0.863787	2	0.825	0.653465	0.729282	3	1	1	1	1,5
5	The Three Musketeers	0.851852	0.901961	0.87619	1	0.794118	0.80198	0.79803	1	0.75	1	0.857143	1
2	USMBA_UIT	0.842466	0.803922	0.822742	2	0.839506	0.673267	0.747253	1	0	0	0	
19	NLPatVCU	0.860927	0.849673	0.855263	1	0.733333	0.653465	0.691099	1	1	0.333333	0.5	2
27	BIT.UA	0.822485	0.908497	0.863354	1	0.773196	0.742574	0.757576	5	0.666667	0.666667	0.666667	1,2
25	Jungfraujoch	0.881481	0.777778	0.826389	1	0.819672	0.49505	0.617284	1	0	0	0	1
24	CLaC	0.738636	0.849673	0.790274	2	0.777778	0.693069	0.732984	3,4	0.5	0.333333	0.4	3,4
26	catalytic	0.734104	0.830065	0.779141	2	0.758621	0.653465	0.702128	2	0	0	0	all
8	DigiLab-UG	0.589641	0.96732	0.732673	3	0.513089	0.970297	0.671233	4	0.75	1	0.857143	2,3,4
1	Trerotola	0.52549	0.875817	0.656863	1	0.470588	0.871287	0.611111	1	0	0	0	1
17	BHAM	0.2875	0.45098	0.351145	1	0.305556	0.326733	0.315789	1	0	0	0	1
11	LasigeBioTM	0.608108	0.294118	0.396476	5	0.407407	0.108911	0.171875	5	0	0	0	all
9	TMU_NLP	0.631579	0.078431	0.139535	2	0.888889	0.079208	0.145455	2	0	0	0	all
12	Elsevier Health Data Science	0.875	0.045752	0.086957	1	0.75	0.059406	0.110092	1	0	0	0	1
20	Orpailleur	0.277778	0.065359	0.10582	2	0.5	0.089109	0.151261	4	0	0	0	all
30	NetPharMed	0.012289	0.052288	0.0199	1	0.023861	0.108911	0.039146	1	0	0	0	1
22	CanSa	0	0	0	1	0	0	0	1	0	0	0	1
mean		0.7631	0.7670	0.7428		0.7141	0.6342	0.6496		0.3648	0.3333	0.3361	
std		0.1994	0.2979	0.2684		0.1963	0.2554	0.2432		0.4599	0.4413	0.4334	
maximum		0.916667	0.9673	0.93		0.8978	0.9703	0.8513		1	1	1	

TABLE VI. MAIN TRACK GRANULAR RESULTS II

#	Team	*DIRECT-REGULATOR*				*ACTIVATOR*				*INHIBITOR*			
		P	***R***	***F1***	***run***	***P***	***R***	***F1***	***run***	***P***	***R***	***F1***	***run***
15	Humboldt	0.764557	0.703963	0.73301	5	0.830247	0.805389	0.817629	5	0.880114	0.880114	0.880114	1
18	NLM-NCBI	0.70362	0.724942	0.714122	4	0.864078	0.799401	0.830482	1	0.873574	0.874405	0.87399	4
13	KU-AZ	0.731458	0.666667	0.697561	1	0.8375	0.802395	0.819572	2	0.873704	0.882017	0.877841	2
7	UTHealth-CCB	0.749296	0.620047	0.678571	2	0.830247	0.805389	0.817629	1	0.864279	0.872502	0.868371	5
21	bibliome	0.710588	0.703963	0.70726	2	0.844037	0.826347	0.835098	5	0.840513	0.872502	0.856209	5
3	CU-UD	0.729268	0.69697	0.712753	4	0.81194	0.814371	0.813154	4	0.856333	0.862036	0.859175	4
29	TTI-COIN	0.73236	0.701632	0.716667	1	0.781977	0.805389	0.79351	3	0.84981	0.850618	0.850214	2
4	good team	0.712531	0.675991	0.69378	5	0.792285	0.799401	0.795827	5	0.836347	0.880114	0.857673	5
23	FSU2021	0.767908	0.624709	0.688946	4	0.808642	0.784431	0.796353	4	0.848309	0.835395	0.841802	4
14	HY-NLP	0.618852	0.703963	0.65867	1	0.745946	0.826347	0.784091	1	0.844844	0.849667	0.847249	2
28	NVhealthNLP	0.730366	0.65035	0.688039	4	0.770393	0.763473	0.766917	4	0.838586	0.835395	0.836988	4
16	HITSZ-ICRC	0.702997	0.601399	0.648241	2	0.827119	0.730539	0.775835	1	0.850354	0.80019	0.82451	4
6	Saama Research	0.694737	0.615385	0.652658	1	0.79375	0.760479	0.776758	1	0.819021	0.843958	0.831303	1
10	Stelios	0.717617	0.645688	0.679755	4	0.737892	0.775449	0.756204	4	0.810029	0.799239	0.804598	5
5	The Three Musketeers	0.651515	0.701632	0.675645	1	0.769939	0.751497	0.760606	1	0.77931	0.860133	0.81773	1
2	USMBA_UIT	0.690608	0.582751	0.632111	2	0.813505	0.757485	0.784496	4	0.833166	0.788773	0.810362	4
19	NLPatVCU	0.700288	0.566434	0.626289	1	0.79322	0.700599	0.744038	1	0.814887	0.791627	0.803089	1
27	BIT.UA	0.649412	0.643357	0.64637	2	0.75841	0.742515	0.750378	2	0.762887	0.84491	0.801806	1
25	Jungfraujoch	0.722222	0.515152	0.601361	1	0.859574	0.60479	0.710018	1	0.834382	0.757374	0.794015	1
24	CLaC	0.589744	0.589744	0.589744	2	0.726444	0.715569	0.720965	1	0.745234	0.818268	0.780045	2
26	catalytic	0.729825	0.484848	0.582633	2	0.698962	0.60479	0.648475	2	0.735238	0.734539	0.734888	1
8	DigiLab-UG	0.375258	0.848485	0.520372	4	0.485531	0.904192	0.631799	4	0.564982	0.893435	0.692223	4
1	Trerotola	0.285959	0.778555	0.418284	1	0.383562	0.838323	0.526316	1	0.286394	0.859182	0.429591	1
17	BHAM	0.239203	0.335664	0.27934	1	0.172606	0.464072	0.251623	1	0.307692	0.449096	0.365184	1
11	LasigeBioTM	0.273381	0.265734	0.269504	1	0.503268	0.230539	0.316222	1	0.397713	0.297812	0.340588	4
9	TMU_NLP	0.517241	0.06993	0.123203	1	0.666667	0.131737	0.22	2	0.637427	0.103711	0.178396	2
12	Elsevier Health Data Science	0.740741	0.04662	0.087719	1	0.807692	0.062874	0.116667	1	0.880597	0.056137	0.105546	1
20	Orpailleur	0.311475	0.044289	0.077551	3	0.15	0.026946	0.045685	2	0.415493	0.056137	0.09891	3
30	NetPharMed	0.03841	0.132867	0.059592	1	0.027668	0.10479	0.043777	1	0.100559	0.239772	0.141692	1
22	CanSa	0	0	0	1	0	0	0	1	0	0	0	1
mean		0.6058	0.5459	0.5557		0.6881	0.6487	0.6495		0.7381	0.7222	0.712	
std		0.1822	0.2102	0.201		0.2063	0.2558	0.2422		0.189	0.2622	0.2441	
maximum		0.775281	0.8485	0.7330		0.8641	0.904	0.8351		0.8953	0.900	0.8801	

TABLE VII. MAIN TRACK GRANULAR RESULTS III

#	Team	*INDIRECT-DOWNREGULATOR*				*INDIRECT-UPREGULATOR*				*PART-OF*			
		P	***R***	***F1***	***run***	***P***	***R***	***F1***	***run***	***P***	***R***	***F1***	***run***
15	Humboldt	0.75	0.848684	0.796296	4	0.776978	0.779783	0.778378	3	0.712062	0.802632	0.754639	1
18	NLM-NCBI	0.701657	0.835526	0.762763	3	0.774194	0.779783	0.776978	4	0.753086	0.802632	0.77707	2
13	KU-AZ	0.758824	0.848684	0.801242	1	0.744186	0.808664	0.775087	3	0.726141	0.767544	0.746269	2
7	UTHealth-CCB	0.755418	0.802632	0.778309	2	0.770073	0.761733	0.76588	2	0.717593	0.679825	0.698198	2
21	bibliome	0.695531	0.819079	0.752266	1	0.707792	0.787004	0.745299	5	0.650519	0.824561	0.727273	5
3	CU-UD	0.723529	0.809211	0.763975	5	0.732026	0.808664	0.768439	5	0.696	0.763158	0.728033	2
29	TTI-COIN	0.724036	0.802632	0.76131	1	0.707006	0.801444	0.751269	3	0.664286	0.815789	0.732283	2
4	good team	0.742671	0.75	0.746318	1	0.710098	0.787004	0.746575	5	0.645283	0.75	0.693712	1
23	FSU2021	0.672775	0.845395	0.749271	1	0.679012	0.794224	0.732113	1	0.641304	0.776316	0.702381	1
14	HY-NLP	0.714689	0.832237	0.768997	1	0.68038	0.776173	0.725126	1	0.599315	0.767544	0.673077	1
28	NVhealthNLP	0.74359	0.763158	0.753247	4	0.760563	0.779783	0.770053	4	0.677725	0.627193	0.651481	3
16	HITSZ-ICRC	0.693694	0.759868	0.725275	4	0.740876	0.732852	0.736842	4	0.670833	0.70614	0.688034	4
6	Saama Research	0.738602	0.799342	0.767773	1	0.690476	0.732852	0.711033	1	0.641921	0.644737	0.643326	1
10	Stelios	0.687879	0.746711	0.716088	5	0.750916	0.740072	0.745455	4	0.722222	0.684211	0.702703	4
5	The Three Musketeers	0.628866	0.802632	0.705202	1	0.61976	0.747292	0.677578	1	0.619658	0.635965	0.627706	1
2	USMBA_UIT	0.665625	0.700658	0.682692	2	0.711679	0.703971	0.707804	4	0.692308	0.513158	0.589421	2
19	NLPatVCU	0.671779	0.720395	0.695238	1	0.676375	0.754513	0.713311	1	0.71066	0.614035	0.658824	1
27	BIT.UA	0.629526	0.743421	0.68175	2	0.701961	0.646209	0.672932	1	0.594771	0.798246	0.681648	1
25	Jungfraujoch	0.725	0.667763	0.695205	1	0.725322	0.610108	0.662745	1	0.674033	0.535088	0.596577	1
24	CLaC	0.633609	0.756579	0.689655	2	0.688	0.620939	0.652751	2	0.509579	0.583333	0.543967	2
26	catalytic	0.788043	0.476974	0.594262	2	0.609848	0.581227	0.595194	1	0.583333	0.644737	0.6125	1
8	DigiLab-UG	0.498141	0.881579	0.63658	3	0.450909	0.895307	0.599758	4	0.431373	0.868421	0.576419	4
1	Trerotola	0.385938	0.8125	0.523305	1	0.331915	0.844765	0.476578	1	0.318731	0.925439	0.474157	1
17	BHAM	0.18724	0.444079	0.263415	1	0.201531	0.285199	0.236173	1	0.179487	0.214912	0.195609	1
11	LasigeBioTM	0.28	0.138158	0.185022	3	0.278049	0.205776	0.236515	4	0.555556	0.109649	0.18315	1
9	TMU_NLP	0.654762	0.180921	0.283505	2	0.592233	0.220217	0.321053	2	0.62766	0.258772	0.36646	2
12	Elsevier Health Data Science	0.6	0.098684	0.169492	1	0.5	0.032491	0.061017	1	0.411765	0.061404	0.10687	1
20	Orpailleur	0.15625	0.016447	0.029762	2	0.191489	0.032491	0.055556	3	0.428571	0.065789	0.114068	4
30	NetPharMed	0.03937	0.131579	0.060606	1	0.016438	0.086643	0.027634	1	0.017133	0.131579	0.030318	1
22	CanSa	0	0	0	1	0	0	0	1	0	0	0	1
mean		0.6103	0.6633	0.6191		0.6157	0.6331	0.6102		0.6059	0.5841	0.5619	
std		0.1805	0.2601	0.2269		0.1868	0.2456	0.2218		0.1599	0.2516	0.2175	
maximum		0.7880	0.8947	0.8012		0.8163	0.8953	0.7784		1	0.8860	0.7771	

TABLE VIII. MAIN TRACK GRANULAR RESULTS IV

#	Team	*PRODUCT-OF*				*SUBSTRATE*				*SUBSTRATE_PRODUCT-OF*			
		P	***R***	***F1***	***run***	***P***	***R***	***F1***	***run***	***P***	***R***	***F1***	***run***
15	Humboldt	0.657277	0.773481	0.71066	5	0.728682	0.673031	0.699752	5	0	0	0	all
18	NLM-NCBI	0.631336	0.756906	0.688442	3	0.772118	0.687351	0.727273	3	0	0	0	3
13	KU-AZ	0.692308	0.696133	0.694215	5	0.737533	0.670644	0.7025	5	0	0	0	all
7	UTHealth-CCB	0.690217	0.701657	0.69589	5	0.775316	0.584726	0.666667	2	0	0	0	all
21	bibliome	0.68617	0.712707	0.699187	5	0.705729	0.646778	0.674969	2	1	0.1	0.181818	2
3	CU-UD	0.651282	0.701657	0.675532	2	0.718919	0.634845	0.674271	3	0	0	0	all
29	TTI-COIN	0.636816	0.707182	0.670157	3	0.683784	0.603819	0.641318	3	0	0	0	3
4	good team	0.578125	0.81768	0.677346	5	0.724324	0.639618	0.679341	3	0	0	0	all
23	FSU2021	0.554656	0.756906	0.640187	1	0.679389	0.637232	0.657635	4	0	0	0	all
14	HY-NLP	0.625	0.635359	0.630137	2	0.623832	0.637232	0.63046	1	0.1	0.1	0.1	2
28	NVhealthNLP	0.696552	0.558011	0.619632	4	0.688474	0.527446	0.597297	3	0	0	0	all
16	HITSZ-ICRC	0.653061	0.707182	0.679045	4	0.700855	0.587112	0.638961	2	0	0	0	all
6	Saama Research	0.589744	0.635359	0.611702	1	0.643045	0.584726	0.6125	1	0	0	0	1
10	Stelios	0.70229	0.508287	0.589744	4	0.591011	0.627685	0.608796	4	0	0	0	all
5	The Three Musketeers	0.557078	0.674033	0.61	1	0.635417	0.582339	0.607721	1	0	0	0	1
2	USMBA_UIT	0.632432	0.646409	0.639344	2	0.679558	0.587112	0.629962	2	0	0	0	all
19	NLPatVCU	0.611429	0.59116	0.601124	1	0.614362	0.551313	0.581132	1	0	0	0	1
27	BIT.UA	0.619318	0.60221	0.610644	2	0.646707	0.515513	0.573705	1	0	0	0	1
25	Jungfraujoch	0.746377	0.569061	0.645768	1	0.710317	0.427208	0.533532	1	0	0	0	1
24	CLaC	0.608187	0.574586	0.590909	1,3,4	0.51134	0.591885	0.548673	2	0	0	0	all
26	catalytic	0.598901	0.60221	0.600551	1	0.652866	0.48926	0.559345	2	0	0	0	all
8	DigiLab-UG	0.279113	0.834254	0.418283	3	0.34714	0.840095	0.491277	4	0	0	0	all
1	Trerotola	0.271881	0.80663	0.406685	1	0.278045	0.828162	0.416317	1	0	0	0	1
17	BHAM	0.115854	0.209945	0.149312	1	0.236324	0.257757	0.246575	1	10	0	0	1
11	LasigeBioTM	0	0	0	all	0.140969	0.076372	0.099071	1	0	0	0	all
9	TMU_NLP	0.183673	0.049724	0.078261	2	0.450549	0.097852	0.160784	2	0	0	0	all
12	Elsevier Health Data Science	0.157895	0.033149	0.054795	1	0.467742	0.069212	0.120582	1	0	0	0	1
20	Orpailleur	0.133333	0.033149	0.053097	3	0.28169	0.047733	0.081633	4	0	0	0	all
30	NetPharMed	0.009533	0.055249	0.01626	1	0.036187	0.195704	0.06108	1	0	0	0	1
22	CanSa	0	0	0	1	0	0	0	1	0	0	0	1
mean		0.5108	0.5576	0.5220		0.5874	0.5055	0.5261		0.0105	0.002	0.0027	
std		0.2154	0.2451	0.2257		0.1903	0.2092	0.2032		0.0975	0.0137	0.02	
maximum		0.7463	0.8343	0.7107		0.7824	0.8449	0.7273		1	0.1	0.1818	